DOCUMENT RESUME

ED 138 637 TM 006 275

AUTHOR Talmage, Harriet; Walberg, Herbert T.
TITLE An Evaluation Research Model for System-Wide Textbook Selection.
PUB DATE [Apr 77]
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Action Research; Classroom Environment; *Elementary Education; *Evaluation Methods; Instructional Materials; Participant Satisfaction; Predictor Variables; Reading Achievement; *Reading Material Selection; School Districts; Teacher Characteristics; *Textbook Evaluation; Textbook Selection

ABSTRACT

One component of an evaluation research model for system-wide selection of curriculum materials is reported: implementation of an evaluation design for obtaining data that permits professional and lay persons to base curriculum materials decisions on a "best fit" principle. The design includes teacher characteristics, learning environment variables, and instructional variables, in addition to materials content analysis and achievement data. After removing the effects of pretest achievement, students' perception of competitiveness in the reading class was a significant cross-validated variable affecting posttest achievement. The findings were instrumental in articulating the system's philosophy of reading and in selecting the "best fitting" reading series. (Author/RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED138637

AN EVALUATION RESEARCH MODEL FOR SYSTEM-WIDE TEXTBOOK SELECTION

Harriet Talmage and Herbert J. Walberg
University of Illinois at Chicago Circle

and

Samuel Nicholas, School District U-46, Elgin, Illinois

1977 AERA Presentation - #14.01

TM006 275

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AN EVALUATION RESEARCH MODEL FOR SYSTEM-WIDE TEXTBOOK SELECTION*[1]

Harriet Talmage and Herbert J. Walberg
University of Illinois at Chicago Circle

and

Samuel Nicholas, School District U-46, Elgin, Illinois

Selection of instructional materials is the Russian roulette of education. Instead of being faced with one out of six loaded chambers, educators have all but possibly one of the six chambers loaded against them. The fatal ammunition is in the form of publishers' blurbs, invalid and unreliable rating scales, biases of outspoken teachers, profunctory acts of ad hoc selection committees, sabre rattling of politically partisan parent groups, and unethical arrangements between publishers and superintendents. With instructional budgets cut to the bone, there is little room for error in the selection of materials. For a large school district, a major textbook series for adoption could easily run from $60,000 to $300,000 or more excluding implementation costs.

The extent of the monetary loss over what may prove an inadequate decision is only outweighed by the effects of an inappropriate curriculum decision. Despite the time, effort, and money school districts put into developing curricula, the basic textbook still remains, in most districts, the arbiter of the curriculum (Talmage, 1972; Talmage, 1971). This holds especially in such subject areas as reading, mathematics, and science. Although there are indications that teacher-developed curriculum guides and teacher-made instructional materials have positive effects on those teachers involved in the development endeavor (Taba, 1962; Beauchamp, 1962), these activities are costly and there is no evidence of transfer value to other teachers. Educators cannot avoid the serious decision making task of selecting appropriate instructional materials.

The evaluation research model herein reported involved an action research design carried out in the naturalistic setting using process and attitude measures as well as achievement measures. Action research as defined by Corey (1952, 1953) involves field investigation by practitioners (teachers, supervisors, and administrators) to improve educational practice. While this approach was in vogue in the 1940's and early 1950's, the methodology associated with action research met with considerable criticism (Hogkinson, 1957; Clifford, 1973) and by the late 1950's few research studies made reference to action research. However, close collaboration between evaluation consultants and school personnel, recent advances in statistical analysis, and greater sophistication in research methodology may begin to overcome the former weaknesses in action research. AERA's special interest group in school-based research and evaluation and works like Good, Biddle, and Brophy (1975) give support to the need for and value of systematic involvement of school personnel in evaluation research.

The investigators were charged by a large metropolitan school district with the task of assisting the staff in designing and implementing an evaluation research study for selecting a basal reading series for district adoption. To avoid a continuation of Russian roulette, the committee assigned the responsibility of recommending a new

*A paper presented at the American Educational Research Association Conference, Division B, New York, 1977.

basal reading series to the School Board, agreed on the need for a large and varied data base from which decisions could be made. In effect, the basal series would become the curriculum.

Previously the staff narrowed selection to four series: Ginn & Company (360), Holt, Rinehard, & Winston, J.B. Lippincott Co., and Scott, Foresman & Co. (Systems). The four series were pilot tested in 12 schools in grades 1,2,3, and 6. A fifth series, Scott, Foresman & Co. (Basic), the then currently adopted series, provided baseline data. A total of 60 classes participated in the study. In addition to reading series, the other predictor variables included teacher characteristics, instructional characteristics, and classroom learning environment. The criterion variable was reading achievement with effects of pre-test scores removed.

Naturalistic Validity

The study needed to account for three sources of invalidity that frequently mar evaluation research: disruption of the naturalistic setting by an obtrusive research design; inappropriate measurement instruments; and faulty implementation of curricula or instructional materials. In designing the study, the district personnel and the investigators held the ongoing educational process as first priority and avoided disturbing the normal instructional activities. If something required modification, it was to be the research design and not the instructional program.

In keeping with the above decision, the SRA Reading Tests, the reading achievement instrument used in the district, were administered as pre and post achievement tests in three of the four grades. For the first grade, the Metropolitan Reading Readiness Test was used as a pretest measure and the SRA as the posttest measure. Inasmuch as the five reading series were purported to emphasize, in varying degree, different aspects of language acquisition and approaches to reading, a SRA Reading Tests item content matrix by reading series was constructed. The matrix determined whether the tests biased the reading achievement scores in the direction of one or another of the basal reading series. The findings indicated a few areas of difference but no overall test bias was observed.

Studies comparing two or more curricula or instructional materials often measure non-events. To use a well-designed reading series for maximum effectiveness, the assumption is that the series must be implemented in keeping with a methodology consistent with the developers' intent (Rosenshine, 1971; Eash, Talmage, and Walberg, 1974; Charters and Jones, 1973). To guard against the possibility of measuring a non-event, data were obtained on teachers' awareness of the design of the reading series they were pilot testing, and by classroom observations of the ongoing instructional activities. Twenty-one items were incorporated in a teacher questionnaire based on the salient design characteristics of the five reading series. Design characteristics were identified as intent, content organization, methodology, and evaluation: a conceptual structure for analyzing instructional materials first developed by Eash (1969). The items yielded an awareness of design score for the series. Significant differences were observed for series design awareness. Teachers pilot testing the Holt reading series obtained the highest mean scores on awareness of design followed by Lippincott and Scott, Foresman Systems teachers. Those using the Scott, Foresman Basic series showed the lowest mean awareness scores despite the fact that these teachers had more years of experience with that series.[2] Holt teachers reported satisfaction with the teachers' manual and the assistance it offered in planning lessons; they also rated their

workshops as more valuable in preparing to use the series than the teachers attending workshops related to the other series which may account for these findings.

Differences in classroom practice were observed that tended to distinguish classess by reading series.[3] Observers found the Scott, Foresman Systems classes and Lippincott classes the most distinctive; the former were more open and individualized, the latter more structured and teacher directed than classes using the other series. The Holt teachers did not frequently use the evaluation component built in to the design of the series; it was reported by teachers to be too difficult to understand and too time consuming. Classes differed in the type of instructional grouping along series lines. The investigators concluded that generally the distinctive design characteristics of each series were being implemented in the respective classes.

Teacher and Class Variables

In addition to series and grade level, the study identified three categories of possible relevant variables for inclusion in the regression equations: teacher characteristics, instructional characteristics, and classroom learning environment. Breakdown of these categories is shown in Table 1. Variables subsumed under teacher characteristics included years in teaching, number of work-shops related to a given series attended, number of reading courses completed for credit, and awareness of design.[4] Under instructional characteristics scores on three variables were obtained from a classroom observation schedule: locus of decision making (i.e., who makes the instruction decisions -- the reading series, the teacher, or the students); student behavior (i.e., manifested by students during reading classes); and the variety and utilization of materials. Scores on five variables categorized under classroom learning environment were obtained by administering the My Class Inventory (Anderson, 1973) to each of the 60 classes. Class means were obtained on the students' perception of Satisfaction, Friction, Competitiveness, Difficulty, and Cohesiveness during their reading lessons.

Analysis

Data were analyzed to yield maximum information for district-wide textbook selection.[5] The class was the unit of study. Pretest reading scores were entered first in all step-wise regression analyses to remove their effects on the posttest reading scores prior to entry of the other variables. Following examination of all main effects, interaction effects were explored to study possible advantages of one series over others given various teaching and instructional characteristics and under various environments. To internally replicate the study and to provide stringent validation of the findings, the classes were sorted into grade levels and a random half of the classes within each grade level was selected for a calibration sample; the remaining half served for cross-validation. Only statistically significant findings across samples are reported.

Observation, interview, teacher characteristics, and series' implementation data not used directly in the regression analyses were analyzed and brought into the textbook selection process. Such data included problems in implementing a series, weaknesses and strengths of each series, and specific difficulties transfer-in students faced when coming from another series.

Table 1. Summary of Variables

DEPENDENT VARIABLE

- Posttest reading achievement

INDEPENDENT VARIABLES

- Pretest reading achievement
- Reading series
  - Ginn 360
  - Holt, Rinehart and Winston
  - Lippincott
  - Scott-Foresman Systems
  - Scott-Foresman Basic
- Teacher characteristics
  - Years of teaching
  - Workshops attended (inservice)
  - Reading courses completed for credit
  - Awareness of design score
  - Attitude toward series
- Instructional Characteristics
  - Locus of instructional decisions
  - Student behavior
  - Instructional materials utilization
- Grade level
  - First grade
  - Second grade
  - Third grade
  - Sixth grade
- Learning environment
  - Satisfaction
  - Friction
  - Competitiveness
  - Difficulty
  - Cohesiveness

Results

In keeping with past research findings, the most significant predictor of reading achievement was the pretest reading scores. After removing pretest effects, only one other variable emerged as a significant predictor of reading achievement. Such variables as series, number of years in teaching, number of workshop sessions attended, number of college reading courses completed for credit, awareness of design characteristics, and the three instructional characteristics did not account significantly for the remaining variance. However, competitiveness in the learning environment proved a crucial variable in predicting posttest reading scores after pre-test effects were removed. The higher the competitiveness perceived by students the lower the reading achievement scores. Competitiveness is characterized by such items in the My Class Inventory as the following:

-- Children often race to see who finish first.

-- Some pupils feel bad when they do not do as well as the others.

-- A few children in my class want to be first all of the time.

Differences among series. A significant interaction was not observed for any of the three instructional variables by reading series on posttest achievement scores, however, significant differences ($F=3.51$; $p < .05$) were obtained on one item on the observation instrument pertaining to the focus of instruction. The item distinguished between reading classes that focussed on a single right answer in contrast with reading classes encouraging inquiry. Mean scores by reading series on this item ranged from high on "encouraging inquiry" to high on "a single right answer" in the following order: Scott, Foresman Systems, Holt, Ginn 360, Lippincott, and Scott, Foresman Basic.

Although no significant differences were obtained for the total mean scores among the five series on the observation instrument on the 24 items, Scott, Foresman Systems had the highest mean scores on 80 percent of the items. The means were calculated on a scale from one to five, one represented high structure, rigid instruction, and little opportunity for student input in instructional decisions; a five represented open or flexible instruction and opportunities for student input. After Scott, Foresman Systems, Holt, Ginn 360, Lippincott, and Scott, Foresman Basic followed in that order from high means scores to lower mean scores.

On frequency of observed type of instruction (total group, small group, tutoring, and diagnosing and evaluating) 41.7 percent of the Lippincott and Scott, Foresman Basic teachers reported using the total class as the major type of instruction. In Scott, Foresman Systems classes, 75 percent of the teachers worked with students in small groups most of the time. Ginn 360 and Holt teachers spent more time in diagnosing and evaluating their students' reading performance than teachers using the other series.

Comparative data were obtained on various aspects of implementing the five reading series. Specific problems related to preparation time, word recognition unique to a series, individualizing instruction, uninteresting materials, and obtaining supplementary materials. A significant X^2 was observed along these categories by series ($X^2=42.67$; $p < .0003$). Preparation time was the major concern of Ginn 360 and Lippincott teachers. Only 18.2 percent of the Scott, Foresman Basic

teachers reported preparation time as too time consuming. However, 63.6 percent of the Scott, Foresman Basic teachers reported that the uninteresting materials created problems. Twenty-seven percent of the Holt teachers reported word recognition problems; this was considerably higher than that mentioned by teachers using the other series.

Holt and Scott, Foresman Systems teachers reported the greatest strength of their respective series was the motivational value of the stories. Over 50 percent of the Ginn, Lippincott, and Scott, Foresman Basic teachers noted skill building as the greatest strength. Weaknesses fell in to several categories. Most noteable were the comments of the Holt and Lippincott teachers; the former cited the testing component as too time consuming; Lippincott teachers cited the over emphasis on auditory learning as a disadvantage to some children.

Transfer-in students from other reading series appeared to have little difficulty entering in to Scott, Foresman Systems and Scott, Foresman Basic. Lippincott teachers reported having to review transfer-in students; Ginn 360 and Holt teachers noted vocabulary transfer-in problems.

Differences among grades. Significant F-values were observed for six items on the observation schedule describing instruction along grade levels. The items included: amount of work displayed ($F=3.59$; $p<.01$); variety of hands-on materials ($F=4.47$; $p<.001$); amount of supplementary materials ($F=3.77$; $p<.01$); teacher-made materials ($F=4.61$; $p<.001$); modality ($F=6.48$; $p<.001$); and pupil attentiveness ($F=4.20$; $p<.001$). The highest means were observed in the first grade classes for hands-on materials, amount of supplementary materials, and varied modalities. Third grade classes were observed to be the most attentive. Second and third grade classes displayed the most student work, and the sixth grade classes used less teacher-made materials than the other grades. No interactions were found for grade level by reading series on these items. While 73.3 percent of the teachers in the first grade reported small groups as the major type of instruction, a greater number of the sixth grade teachers than those in the other grades reported tutoring as an important type of instruction.

Selecting the Basal Reading Series

The selection committee had to find a way to sort out the information. Although reading series, per se, did not affect achievement differentially, nor did the degree of teachers' awareness of the instructional design of a series, we cannot infer from this that poorly developed materials are as effective or ineffective in enhancing achievement as are well designed instructional materials. The four series selected by the school district for pilot testing were initially carefully screened from among a large pool of reading materials. These four series represented basal series that fulfilled criteria of well designed instructional materials: materials whose stated objectives, selection and organization of content, supporting methodology,

and evaluation procedures were consistent among these design constructs. The findings do suggest that once materials are screened for design congruence and other selection factors (e.g., community acceptance, biases, cost, graphics), which series is selected among those having the essential qualities will not differentially affect reading achievement. If reading series is removed as a selection variable, given that all the series were equally well designed, the findings suggested that the selection committee explore the implications of a competitive learning environment in reading achievement. Competition was perceived by many teachers, parents, and School Board members as supportive of learning. Reared themselves to compete for grades and having successfully achieved academically in a competitive environment, many teachers and Board members were hesitant to turn their attention from textbook selection to examining the District's philosophy and the educational-psychological implications of a competitive learning environment.

Children who are slow in reading will be obviously behind in the eyes of their classmates in the traditional classroom. Those at the top of the class in reading must constantly be on guard to maintain their positions if there are high levels of competition. Similarly those in the middle of the class may be under pressure if they are always being compared with one another.

None of the above means, of course, that competition is always detrimental to learning. But reading is a very serious business in the eyes of the child, his classmates, teacher, and parents; reading pervades all areas of the school curriculum including science and mathematics. To be worst or even second best in a highly important, competitive situation for a child in the early and middle grades may be defeating. The all too familiar consequences of failure need no recounting.

The role of competitiveness in reading achievement has implications for instruction, certainly for the need to share with the students the progress they are making without the pain and distraction of unfair or invidious comparisons of their work with that of others. Other aspects of instruction that tend to reduce competitiveness require further study, such as grouping arrangements (other than by ability), evaluation approaches, and pacing.

From classroom observations, interviews, and questionnaire data, teachers that best exemplify types of individualization that reduce competitiveness are not necessarily more experienced, nor technically better trained, nor even more aware of the design of reading materials. Further implications can be drawn from these findings relative to both preservice and inservice teacher education.

The question about selection of basal reading materials cannot be separated from instructional implementation. A basal reading series selection committee must ask two questions: (1) Which basal series can be most readily implemented in a way that supports a noncompetitive learning environment? (2) What support needs to be provided the classroom teacher to ensure development of non-competitive instructional approaches?

The focus of selection shifted from a single best series to the "best fit" between the reading series and the District's educational philosophy. On the recommendation of the selection committee, the School Board approved a single reading series. The decision was based on the need to put the effort and money into implementation: to

provide teachers with the support needed to use the materials in non-competitive ways. This meant changes in inservice programs, in supplementary materials, and in classroom practices.

NOTES

1. The investigators are indebted to Barbara Furlong, Reading Coordinator, and the faculty and staff of School District U-46, Elgin, Illinois; Sue Pinzur Rasher, Educational Specialist, UICC; and Roberta Talmage. The research was supported in part by Illinois School District U-46, The Office of Evaluation Research at the University of Illinois at Chicago Circle, and the Urban Education Research Program (UICC).

2. Scott, Foresman Basic teachers attended fewer inservice workshops in preparation for teaching with the series than teachers using the other four series. It is a truism of instruction that the longer instructional materials are used, the more adaptations teachers feel free to make, hence the further they move from the original design of a series.

3. A committee composed of teachers and administrators selected from the District's Language Arts Task Force and the investigators modified the Kunkel-McElhinney observation instrument for this study.

4. Data on other teacher characteristics were also collected. Only these four variables, however, were included in regression equations inasmuch as results from Chi Square and F-tests indicated significant differences by series on these variables.

5. For further discussion of statistical analyses see H. J. Walberg and H. Talmage, "City-Wide Action Research on Reading Achievement," Office of Evaluation Research, UICC, 1976.

REFERENCES

Anderson, G.J. The Assessment of Learning Environments: A Manual for the Learning Environment Inventory and the My Class Inventory (2nd ed.). Halifax, Nova Scotia: Atlantic Institute of Education, 1973.

Beauchamp, G.A. Curriculum theory applied to urban schools. In Chandler, B.J., Stiles, L.J., and Kitsuse, J.I., (Eds.). Education in Urban Society. New York: Dodd, Mead & Company, Inc., 1962.

Charters, W.W., Jr., and Jones, E. On the risk of appraising non-events in program evaluation. Educational Researcher, 1973, 2, 5-7.

Clifford, G.J. A history of the impact of research on teaching. In Travers, R.M.W., (Ed.). Second Handbook of Research on Teaching. Chicago: Rand McNally & Co., 1973.

Corey, S.M. Action Research to Improve School Practices. New York: Bureau of Publications, Teachers College, Columbia University, 1953.

Eash, M.J. Assessing curriculum materials: A preliminary instrument. EPIE Educational Products Report, 1969, 2:5, 18-24.

Eash, J., Talmage, H., and Walberg, H.J. Evaluating instructional systems. EPIE Educational Products Report, 1974, No. 58, 64pp.

Good, T.L., Biddle, B.J., and Brophy, J.E. Teachers Make a Difference. New York: Holt, Rinehart and Winston, Inc., 1975.

Hogkinson, H.L. Action Research: A Critique. Journal of Educational Sociology, 1957, 31, 137-153.

Rosenshine, B. New directions for research in teaching. How Teachers Make a Difference. U.S. Office of Education, Bureau of Educational Personnel Department. Washington, D.C.: U.S. Government Printing Office, 1971.

Taba, H. Curriculum Development: Theory and Practice. New York: Harcourt, Brace and World, Inc., 1962.

Talmage, H. The textbook as arbiter of curriculum and instruction. The Elementary School Journal, 1972, 73:1, 20-25.

Talmage, Harriet. The use of textbooks in individualized instruction. Curriculum Advisory Services Quarterly, 1971, 10:2, 57-60.